Journal of the International Neuropsychological Society (2000), **6**, 608–619.

Simulation of cognitive disturbances by a dynamic threshold semantic neural network

AMIR B. GEVA[1] AND AVI PELED[2]

[1]Electrical and Computer Engineering Department, Ben-Gurion University of the Negev, Beer Sheva, Israel
[2]Department of Forensic Psychiatry, Shaa'r Menashe Mental Health Center, Israel; Technion-Israel Institute of Technology, Haifa, Israel

(Received May 15, 1999; Revised October 1, 1999; Accepted October 8, 1999)

Abstract

A neural network model with dynamic thresholds, asymmetric connections, and clustered memories simulates spread activation that is hypothesized for semantic networks in the brain. By altering the parameters of the dynamic threshold a large range of disturbances can be generated in the model. These disturbances show metaphorical resemblance to certain general clinical descriptions of mental disturbances found in psychiatric patients engaged in various cognitive tasks. Even though the model is highly theoretical and metaphoric, it may help to gain certain insights into the relation between alterations of certain neural parameters, for example, thresholds and connectivity, and clinical symptoms in patients. (*JINS*, 2000, *6*, 608–619.)

Keywords: Dynamic neural networks, Semantic networks, Cognitive task, Working memory, Concrete and abstract concepts representation, Schizophrenia, Thought disorders, Fuzzy clustering

INTRODUCTION

Recently neural network models have been used to simulate many normal and pathological mental functions. The simulation of normal cognitive functions such as learning, memory, recognition, and categorization has provided impressive results (Hinton, 1981, 1986; Rumelhart & McClelland, 1986). Simulations of cognitive disorders such as disturbances of memory activation and associations have also shown considerable achievements (Cohen & Servan-Schreiber, 1992, 1993; Hermann et al., 1993; Hinton & Shallice, 1991; Hoffman, 1987, 1992; Hoffman & Dobscha, 1989; Hoffman et al., 1994; Servan-Schreiber et al., 1996). Typically simulations of cognitive disturbances involve a decline in the performance of a neural computation task as a direct consequence of a change in one of the networks' parameters, usually the threshold function. However, both the neuroscience literature and clinical experience suggest that brain functions are more complicated (Tucker, 1998; Van-Praag, 1997; Wilson, 1993). Neuroscience teaches us that many neuronal parameters (e.g., threshold, connectivity, and inputs) continuously change over time, and that a "balanced" interaction among many dynamic changes occurs during the normal functioning of the brain system (Globus, 1992; King, 1991). Clinical experience indicates an extraordinary variability (or spectrum) in the manifestation of cognitive disturbances (Spitzer & Williams, 1995; Tucker, 1998; Wilson, 1993).

It seems that while simple neural network models (i.e., models of fixed inputs, connections, and threshold alterations) are sufficient for simulating circumscribed mental disturbances, more complex models are required for simulating the complex variety (or spectrum manifestations) of mental disorders. Thus, increasing the intricacy of the model to approximate some of the complexities in the brain may serve to illustrate certain spectrum manifestations in mental disturbances not explained otherwise. In this work, the application of (1) dynamic threshold function, (2) asymmetric connections, (3) clustering of memory patterns, and (4) "internal inputs" offer the necessary complexity to simulate variability and spectrum phenomena in mental disturbances. To demonstrate that certain common neural mechanisms can generate a wide variability in different cognitive functions, four different mental functions and their relevant tasks are chosen. The model simulates the disturbances typically described in the psychiatric literature for

Reprint requests to: Dr. Amir B. Geva, Electrical and Computer Engineering Department, Ben-Gurion University of the Negev, P.O.B. 653, Beer-Sheva 84105, Israel. E-mail: geva@ee.bgu.ac.il

https://doi.org/10.1017/S1355617700655108 Published online by Cambridge University Press

each of these cognitive tasks. The cognitive functions and their corresponding tasks are summarized in Table 1.

Normal and Abnormal Cognitive Functions

Normal thinking tends to be linear, with each idea following the previous one in a relatively ordered association. The thinking process is *goal-directed* and appropriate to the relevant information at hand (e.g., answering a question in the course of a discussion). As such, the associations are related to one another, to the incoming information, and to a particular outcome or response.

With *loosening of association* patients clearly demonstrate a breakdown in linear and goal-directed thought. Ideas are disconnected or obliquely related, sometimes to the extent that the listener cannot understand the speech of the speaker. Consequently, the patient's thinking may be entirely unrelated to conversational topics, or what may start out being related quickly gets off track (i.e., derailed). In contrast, *poverty of thought* involves few associations at all, as if the patient's thoughts are fixated to few ideas. As such, the patient may manifest repetitive, stereotyped, and perseverative thinking. Poverty of thought may be under- or overproductive, but in either case it reflects thinking that is neither linear nor goal-directed. For purposes of the model in this article, we emphasize these two types of thought disorder.

In the *lexical decision task* participants are asked to identify words from nonwords that are flashed on a screen. If, in addition, prior to the word stimulus, a meaningful associated word is presented, the task is performed more quickly and with fewer errors. This phenomenon has been termed "priming," and was found to be significantly impaired in patients who suffer from thought disorders (Manschreck et al., 1988). Indirect priming is the condition where the associated word is indirectly relevant to the word stimulus; for example, chalk is associated with milk through the intermediate concept of white. Indirect priming has been found to be even more impaired in thought-disordered patients (Manschreck et al., 1988). For the purpose of the model in this work we emphasize the associations between the word stimuli. Nonassociated word stimuli take longer to recognize and correlate, with more errors in the lexical decision task.

Working memory is a short-term memory activated and held in mind to monitor and respond correctly to stimuli (Goldman-Rakic, 1994). For example, in delay response tasks the rule for responding to the task is held in mind during the delay period and guides the responses. In the *object alternation task* the participant is asked to choose between two objects according to a predetermined rule: for example to choose between a blue and a red object by alternating the color each time. If the participant performing the task has difficulty maintaining the alternation rule over the delay period he will make errors by choosing the same color twice in a row, or by choosing in an irregular manner violating the rule.

In the *sorting task* the participant is asked to sort a deck of cards bearing stimuli that vary in number, color, and shape. As each card in the stack comes up the participant has to match it to a set of reference cards. Matching is based on one dimension (e.g., color) that is arbitrarily selected by the experimenter. The experimenter then informs the participant whether he or she is right or wrong, and the patient attempts to get as many correct matches as possible. After the patient achieves a specific number of consecutive correct matches, the sorting principle is shifted without warning, and the patient must modify his responses accordingly. Failure in the task occurs when the patient fails to create or maintain the principle that guides the sorting; this is sometimes termed the loss of a set. Other errors frequently documented for patients suffering from mental disturbances involve perseverations, where the patients fail to change set, and thus continue to respond according to the previous principle (Berman et al., 1995). For the purpose of the model in this work, the loss of set and the inability to shift to a new set are emphasized.

Modeling Disturbances in Cognitive Functions by Neural Networks

According to network models of the mental lexicon, semantic information is represented as nodes in a neural network (Collins & Loftus, 1975). These nodes become activated for a short period of time when used for speech production or understanding (Collins & Loftus, 1975; Collins & Quillian, 1970). Based on this description, thought is sometimes conceptualized as *sequences of memory retrievals* (Collins &

Table 1. External (E_i) and internal (I_i) inputs for the entire cognitive task simulations

Cognitive function	Cognitive task	External input (E_i)	Internal input (I_i)
Organized speech	Goal-directed associations	1, 2, 3–* 1, 4, 7	None
Priming	Lexical decision task	1, 2–* 1, 7	None
Working memory	Object alternation task	1, 2, 7, 8–* 1, 2, 7, 8	1, 2–* 1, 7
Abstraction and categorization	Sorting task	3, 4, 7–* 1, 6, 8–* 1, 5, 9–*	1, 2, 3–*
		3, 4, 8–* 1, 2, 7–*1, 4, 8	None

*Change of input.

https://doi.org/10.1017/S1355617700655108 Published online by Cambridge University Press

Loftus, 1975), or activation of *semantic concepts* (Neely, 1991), which once activated in a *coherent sequence* guide speech formation (Desse, 1987; Maher et al., 1983). This description is in line with Quillian's (Collins & Loftus, 1975) basic theory of the *semantic network model*. In his semantic system, concepts are represented as *nodes* and *connections* of a network structure. *Concepts*, which are represented by the nodes in the semantic network, can be represented by *attractors*, or *memories* of a *neural network* (NN) model.

Typically a simulation of a "pathology" in a neural network model involves altering the threshold function (i.e., the bias or the gain; Figure 1) or the connection values of synaptic strength (Cohen & Servan-Schreiber, 1993; Herz et al., 1991; Hinton & Shallice, 1991; Hoffman, 1992; Rumelhart & McClelland, 1986; Servan-Schreiber et al., 1996). However, within the simulation all parameters are maintained constant. Pathology is simulated by disturbances in the regular memory activation process or, in other words, by disturbances in the process of convergence to the relevant attractor (Hoffman, 1992). The model can activate a memory (i.e., normality) or fail to activate a memory (i.e., pathology); thus within the pathological state there is only one limited type of failure, the inability to converge and activate the memory. However as mentioned above, in psychiatric clinical experience a multitude of abnormal cognitive patterns exist. For example, memories can be activated correctly but not in the "right" order. Pathology of memory activations (e.g., concepts in speech or response decisions) can be disturbed in a variety of forms, from mild irregularities to total confusion (e.g., from tangentiality and derailment to loosening of associations and word salad (Andreasen, 1997; Andreasen & Olsen, 1982)). A memory can be activated correctly, but then activated repeatedly again and again in the wrong settings, thus becoming a disturbance or an abnormality (i.e., perseveration (Andreasen & Olsen, 1982)).

Hoffman (1987, 1992; Hoffman & Dobscha, 1989) presented a Hopfield-like *attractor neural network* (ANN) model that becomes overloaded due to synaptic deletion, causing spurious memories to emerge. He suggested that this model could explain mechanisms underlying some schizophrenic symptoms such as hallucinations and delusions. Cohen and Servan-Schreiber (1992, 1993; Servan-Schreiber et al., 1996) presented connectionist feed-forward back-propagation networks that are able to simulate normal and schizophrenic performance in several attention and language-related tasks. Horn and Ruppin (Hermann et al., 1993) investigated the effect of synaptic compensation on the dynamic behavior of the ANN. Spontaneous, stimulus-independent retrieval of stored patterns that occur due to compensation are used as a framework to simulate the underlying mechanisms of loosening of associations.

A *dynamic threshold neural network* (DTNN) has been applied to a variety of findings in semantic activation and memory effects (Horn & Usher, 1989, 1990). Hermann and coworkers (Hermann et al., 1993) have used the DTNN to simulate serial concept activation. By applying dynamic threshold activity (Horn & Usher, 1989, 1990), transient serial memory retrievals are accomplished by the DTNN. Such models have been recently proposed in the neural network literature on the basis of synaptic delays (Abbott et al., 1997; Wickliffe & Warren, 1997), neural adaptation (Bliss & Gardner-Medwin, 1973; Stone et al., 1998), or slow inhibition (Abbott et al., 1997; McNaughton, 1982). Varying the noise level at the unit level in the DTNN model, Hermann et al. (1993) demonstrate the well-documented semantic priming phenomena (Manschreck et al., 1988; Neely, 1977). They have also used the *same* DTNN to simulate some neuroanatomical and neuropsychological findings in Alzheimer's disease (AD). Random synaptic deletion of the DTNN units resulted in a reduction of transitions between memory retrievals. It was suggested that synaptic deletion simulates neuronal loss in AD, and reduction of transitions between memory retrievals simulates the decrease of semantic memory in neuropsychological tests of such patients.

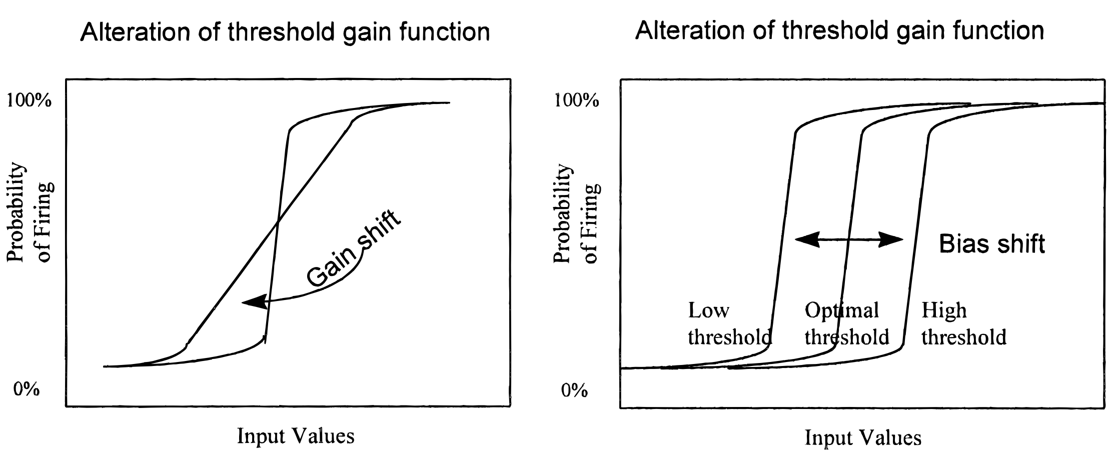

Fig. 1. The input–output threshold of each neuron is modeled by a sigmoid function. Axis *X* indicates the input value and axis *Y* the "firing" probability of the unit. Zero stands for "no firing" and 1 stands for maximal firing rate. Threshold manipulations include alterations of gain (left graph) or alterations of bias (right graph) that is used in the present work.

https://doi.org/10.1017/S1355617700655108 Published online by Cambridge University Press

Our interest in sequences of memory retrievals, or coherent activations of semantic concepts, in a *semantic network model*, leads us to continue and study the *threshold dynamics* of the DTNN. This paper presents a new representation of a semantic network model. The random memory patterns of the DTNN are clustered into classes of near (Hamming distance) patterns, and asymmetric connections are assigned between the units of the DTNN, which represent memory patterns of the same class. Each memory pattern can represent a concrete concept or a node in the semantic network, while each class of associated concepts or "near nodes" in the semantic network can represent an abstract concept. Abstract concepts are represented in the network by activation of a linear combination of a connected group of memory patterns that represent the concrete concepts that constitute it. Using this *dynamic threshold semantic neural network* (DTSNN) model, we attempt to simulate the four cognitive functions described above.

METHODS

The DTSNN Model

The network is constructed from N neurons, characterized by two-valued variables $S_i \in \{-1,1\}$ corresponding to non-active and active states. Each neuron is subject to a dynamic threshold variable θ_i (Horn & Usher, 1989, 1990). Neurons are connected through synaptic strength values called weights W_{ij} (the connection between neuron i and neuron j). The postsynaptic potential, h_i, is the weighted sum of the states of the neurons from which it receives connections:

$$h_i(t) = \sum_{j=1}^{N} w_{ji} S_j(t). \tag{1}$$

The dynamic equation for the output of the units is given by:

$$S_i(t+1) = F_T[h_i(t) - \theta_i(t) + E_i(t) + I_i(t)], \tag{2}$$

where the neuron state, F_T, is a threshold dependent stochastic function:

$$F_T(x) = \begin{cases} +1 \text{ with probability } \left(1 + e^{-\frac{2\cdot x}{T}}\right)^{-1} \\ -1 \text{ with probability } \left(1 + e^{\frac{2\cdot x}{T}}\right)^{-1} \end{cases}, \tag{3}$$

where T represents the "slope" of threshold (i.e., the "temperature" of the model; Figure 1).

The original *external* input to the network $Ei(t)$ consists of a linear combination of subset of memory patterns:

$$E_i(t) = \epsilon \sum_{k\in U} m_i^k. \tag{4}$$

This value represents the composite input from the external world to the model, where the parameter ϵ is the relative weight of the external inputs and U is a set of indexes of the memory patterns in the composite external input.

In addition, we choose to add another composite *internal* input to the network:

$$I_i(t) = \delta \sum_{k\in V} m_i^k, \tag{5}$$

where δ is the relative weight of the combined internal inputs and V is a set of indexes of the memory patterns in the composite internal input. One can regard this value as representing the short-time working memory of the model, which is created by fast learning or adaptation of the model to new situations. At this stage we did not try to model the dynamic behavior of the internal input, which will require further investigation.

The *threshold dynamic* is given by Horn and Usher (1989, 1990):

$$\theta_i(t+1) = \theta_i(t)/c + bS_i(t+1). \tag{6}$$

According to this equation, while a neuron is active, its dynamic threshold $\theta_i(t)$ increases asymptotically (i.e., integration of the neuron activity) to the value

$$\theta_{\max} = cb/(c-1) \tag{7}$$

and deactivates the corresponding neuron. The parameters b and c ($c > 1$) represent the rate of increase and decay of the dynamic threshold respectively. The dynamic threshold provides the motion of the network in the memory-concept space; neurons that are active for a relatively large time are deactivated, and the network's state evolves into a new pattern.

The network learns its memory patterns by Hebbian rule with asymmetric connection between associated patterns (Herz et al., 1991). Assignment of asymmetric connections between successive memory patterns causes the next convergence to occur into the next memory pattern in a sequence of associative memories. The weight between the i-th and the j-th neuron is:

$$W_{ij} = \frac{1}{N}\sum_{k=1}^{P} m_i^k m_j^k + \lambda \cdot \frac{1}{N}\sum_{l=1}^{L} \sum_{\substack{k,q\in G_l \\ k\neq q}} m_i^k m_j^q, \tag{8}$$

where P is the number of memory patterns, L is the number of categories, G_l is a set of indexes of the memory patterns in the l-th category (i.e., G_l contains the indexes of the memory patterns of the l-th category), and the parameter λ is the "strength" of the asymmetric connections.

Simulation of Cognitive Tasks by the DTSNN

The NN memory patterns represent the predetermined semantic concrete concepts acquired by learning; each mem-

https://doi.org/10.1017/S1355617700655108 Published online by Cambridge University Press

ory pattern represents a node of the semantic network. Semantic activation is simulated by memory retrieval, namely convergence into one of the memory patterns of DTSNN. By adding asymmetric connections between the units of the DTSNN, we simulate connections, or associations, between near concepts or nodes in the semantic network. Abstract concepts are then represented in the network by a linear combination of a *connected group* of memory patterns; these are the concrete concepts that constitute each abstract concept.

All the cognitive tasks that were chosen to be simulated by the model are performed by the same neural network model; only the input modalities vary between the different simulations. In the current realization, the network is constructed from 600 neurons, and learns (Equation 8) nine random memory patterns, each representing a concrete concept. The nine random memory patterns are *clustered* into three clusters with three patterns, each according to their Hamming distance. We label the first cluster G_1, and its memory patterns 1, 2, and 3; the second cluster G_2, and its memory patterns 4, 5, and 6; and the third cluster G_3, and its memory patterns 7, 8, and 9. The three memory patterns of each cluster are associated by asymmetric connections (Equation 8). The subsets are chosen to represent three different categories or abstract concepts (Figure 2). A correlation value of .7 between the network state and one of the memory patterns is chosen to define the memory retrieval state (all memory patterns are partially activated continuously).

The *external inputs* (E_i) are constructed from a linear combination of the U set of memory patterns (Equation 4; Table 1), and the *internal inputs* (I_i) are constructed from a linear combination of the V set of memory patterns (Equation 5; Table 1). One can view the external input as a mixture of information coming from the environment, in this case a mixture of concepts embedded in words or sentences, while the internal inputs can represent an internal rule, which is acquired by the short-term memory system. The model can organize an external input information into sequences of ordered input-dependent memory retrievals using the internal input as a guiding rule for its behavior. We assume that choosing the input patterns for the retrieval of memories from the same category may simulate the activation of associated semantic concepts, whereas the retrieval of memories from different categories may simulate the activation of nonassociated semantic concepts.

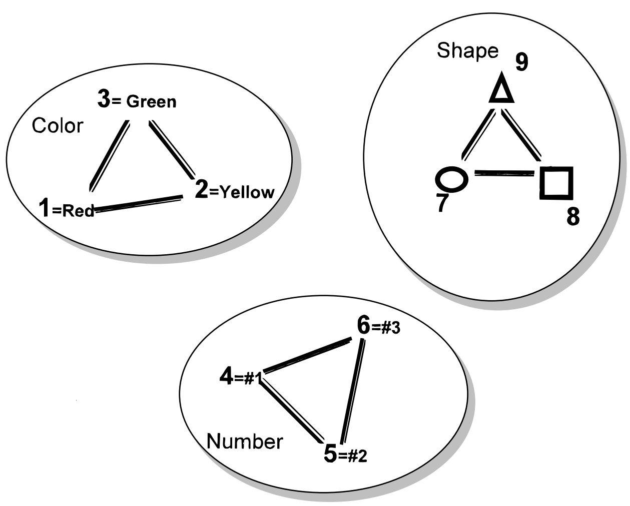

Fig. 2. Concepts and abstract concept are represented in the model by memory patterns that are clustered into categories. Memories 1, 2, and 3 represent colors *red*, *yellow*, and *green*. Memories 4, 5, and 6 represent numbers *1*, *2*, and *3*. Memories 7, 8, and 9 represent shapes *circle*, *square*, and *triangle*. These representations are used for the simulation of the various mental tasks.

Semantic priming is simulated with associated input patterns (small Hamming distance and asymmetric connections), which represent neighboring nodes in the semantic network or concepts from the same category (e.g., *1 = red* and *2 = yellow*, both colors; Figure 1). Nonpriming is simulated by presenting input patterns from different categories (large Hamming distance and no asymmetric connections), which represent far nodes in the semantic network (*1 = red*, *7 = circle*). Priming phenomena are measured by the average time between consecutive convergences into the relevant pairs of memories, which is achieved for each input representation. Frequent convergence into pairs of patterns from the same category, in contrast to fewer convergences into pairs from different categories, may simulate the priming effect.

For the simulation of the object alternation task the external input is set to a linear combination of memories *1* (*red*), *2* (*yellow*), *7* (*circle*), and *8* (*square*). This external input represents the stimulus of this task, which is composed of circles and squares alternately colored in red and yellow. The internal inputs are *1* (*red*) and *2* (*yellow*), which represent the rule of *color alteration*, for the first half of the simulation, and *1* (*yellow*) and *7* (*square*), which represent the rule of *color–shape alteration*, for the second part of the simulation.

For modeling a sorting task by the DTSNN, we assume that each feature on the cards (form, color, and number) can be represented by a memory pattern; for example, *two-red-squares* are represented as memory patterns *1*, *6*, and *8* (*1 = red color*, *6 = number two*, *8 = square form*; Figure 1). Next, we assume that an abstract concept, like *color* can be represented as a linear combination of all the concrete concepts that it stands for, such as all the colors (*1 = red*, *2 = yellow*, *3 = green*). Performing a sorting test is simulated by the DTSNN by presenting an external input (Equation 4) composed of memory patterns that specify the card to sort (e.g., memory patterns *1–6–8* for *red-two-squares*), together with an internal input (Equation 5) comprised of the set of memory patterns that represent the abstract concept, or rule, to follow (e.g., memory patterns *1–2–3* for color). These two sets of memory patterns (i.e., U and V) are matched by the system, and since one memory pattern will always appear twice (in this case memory pattern *1*), it will have a higher probability of being activated. In the presentation we will not deal with the very relevant problem of how does the system acquire (or does not acquire) a new abstract concept. At this stage, we prefer to avoid this problem by assuming that the participant acquires a new abstract concept as soon as he gets a positive answer from the tester. By this we acknowledge the fact that the model cannot simulate errors that are caused by interference in abstract concept acquisition.

External and internal inputs (i.e., E_i and I_i) for each task simulation are presented in Table 1. To summarize, input-dependent sequences of memory retrievals, or coherent activations of semantic concepts, are chosen for the simulation of normal thought processes that guide coherent goal-directed speech. The activation of different associated pairs of semantic concepts is used for the simulation of priming phenomena. Finally, linear combinations of a connected group of memory patterns are suggested to simulate categorization or abstract concepts, which are relevant for decision-making in a sorting task. The model's performance for all the different tasks is checked with the same set of dynamic threshold parameters.

Challenging the Network Performance for Simulating Interferences in Cognitive Tasks

The dynamic threshold parameters b, c, and $\theta_{\max}$, which locally control the activity level of each unit, have a global organizing capacity essential for the network's computational abilities as a whole system. We chose to study the effect of changes in these relevant dynamic threshold parameters on the performance of each of the four cognitive tasks that are simulated by the DTSNN. The other parameters of the network were set at constant values of $T = 0.4$ (the threshold gain; Equation 3 and Figure 1, left), $\epsilon = 0.25$ (the relative weight of the external inputs; Equation 4), $\delta = 0.25$ (the relative weight of the internal inputs; Equation 4) and $\lambda = 0.1$ (the strength of the asymmetric connections; Equation 8). It is important to emphasize that, although these latest parameters are relevant to the network dynamics, we prefer in this research to concentrate on the dynamic threshold parameters. The dynamic threshold parameters *space* of b, c, and $\theta_{\max}$ are presented in Figure 3. The model performances were tested for a wide range of dynamic threshold parameter values. Optimal performances of all four task simulations (semantic sequencing, priming, working memory, and sorting task) occur when the threshold dynamic parameters are around the values of $b = 0.1$, $c = 1.125$, and $\theta_{\max} = 0.9$. Any significant alteration of these values causes some major breakdown or disturbances in the performance of the model, thus causing specific sets of typical disturbances in each of the four simulations. After testing the model's activity for a large number of points in the threshold parameter space, the five points in Table 2 have been chosen to demonstrate the model's normal and disturbed activity.

SIMULATION RESULTS

The five points of interest (Table 2) are described as five different combinations of threshold dynamics:

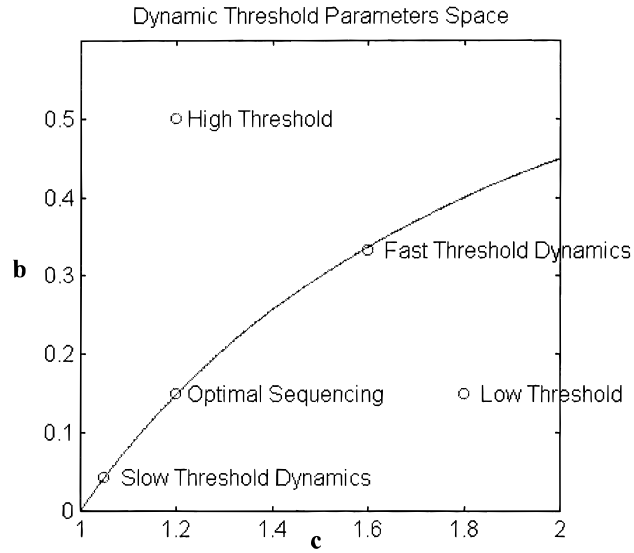

Fig. 3. The dynamic threshold parameters space. Each point in this space represents a different combination of dynamic threshold values, which result in different behavior of the system. The plotted line correspond to b and c values resultant in constant $\theta_{\max} = 0.9$ (Equation 7).

1. The point of optimal threshold variables is the combination of threshold parameters where the model simulates well-organized sequences of spread activation.
2. The point of high and fast threshold dynamics.
3. The point of low and slow threshold.
4. The point of fast, but optimal, levels of threshold.
5. The point of slow, but optimal, levels of threshold.

The last four points describe dynamic threshold combinations that generate various irregularities that make part of an entire range of disturbances in the model.

The simulations results are presented in Figures 4 to 7, each corresponding to a task according to Table 1 (i.e., the simulation in Figure 4 corresponds to thinking, the simulation in Figure 5 corresponds to priming, the simulation in Figure 6 corresponds to object alternation, and the simulation in Figure 7 corresponds to the sorting task). Each graph within each figure is a different simulation run under a different combination of threshold parameters (Table 2).

Axis X for each graph indicates *time*, where time is measured as the number of iterations in the model, that is, the number of times in which all the units are updated. Axis Y for each graph indicates the correlation between the state of the model at the given time and the memory patterns embedded in (or learned by) the system. For example, a curve marked with the number *2* that reaches the level of 1 indicates high correlation between the state of the network and memory pattern number *2*. In other words, it indicates that the network converged to memory number *2* and has activated the information represented by this network pattern

https://doi.org/10.1017/S1355617700655108 Published online by Cambridge University Press

Table 2. Threshold dynamic parameters (b, c, and θ_{max}) that were used for the simulation of the disturbances in the different cognitive tasks

Threshold dynamic variables	b	c	θ_{max}
a) The point of optimal threshold	.1	1.125	0.9
b) The point of high and fast threshold dynamics	.15	1.02	7.65
c) The point of slow and low levels of threshold	.15	1.5	0.45
d) The point of fast but optimal levels of threshold	.22	1.324	0.9
e) The point of slow but optimal levels of threshold	.0089	1.01	0.9

(see introduction above). Although numbers can be the same for different simulations (e.g., thinking, priming, and sorting), in each simulation they mean different information. For example, in thinking the memories are concepts and ideas in the stream of speech, while in the sorting task they are decisions and responses to the card stimuli.

In all simulations the input is changed in the middle of the simulation (vertical line in the figures). In the first three simulations (Figures 4, 5, and 6, of thinking, priming, and object alteration) inputs in the second half of the simulations originate from patterns that are clustered in different categories (see Figure 2 and Table 1). In the last simulation (the sorting task in Figure 7), the vertical line indicates when the rule of deciding according to color is stopped. For each simulation the input is presented at the top and the graph indicates the convergences over time. If there is a rule or a guiding principle it is presented in each figure under the input specifications.

Simulation of Normal Cognitive Performances on the Tasks

The point of optimal sequence activity in the model simulates the normal performance on the four different tasks (Figures 4a, 5a, 6a, and 7a). An ordered sequence of activation of memories *1*, *2*, and *3* for the first half of the iterations, and *1*, *4*, and *7* for the second half (Figure 4a) is considered to simulate linear, goal-oriented (i.e., input-dependent) thinking. Notice that in the first half of the simulation, memories from the same category are activated (i.e., *1*, *2*, and *3* are color categories in Figure 3), while in the other half of the simulation memories from separate categories are activated.

The activation of memory patterns *1* and *2* simulates the two word-stimuli for the lexical decision task (Figure 5a). Ordered output sequences of memories *1* and *2* simulate the normal priming for the lexical decision task (Figure 5a). Activation of memories that are part of different clusters, that

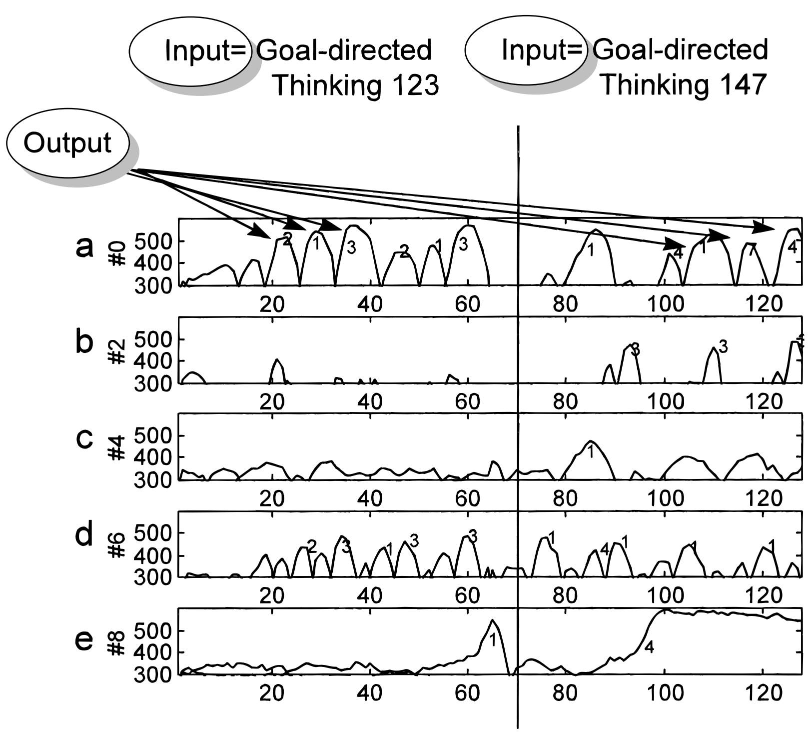

Fig. 4. Each graph *a*, *b*, *c*, *d*, and *e* is a different simulation run under a different combination of threshold parameters presented in the previous Figure 3. Axis *X* indicates *time*, where time is measured as the number of iterations in the model, i.e., the number of times in which all the units are updated. Axis *Y* indicates the correlation between the state of the model at the given time and the memory patterns embedded in (or learned by) the system. The input is presented as a mixture of memory patterns—see also Table 2. The input is changed in the middle of the simulation (vertical line). (a) Simulation of goal directed input-dependent thinking. (b) Simulation of loosening of associations. (c) Simulation of perseverative ideation. (d,e) Simulations of fast and slow ideations, respectively.

https://doi.org/10.1017/S1355617700655108 Published online by Cambridge University Press

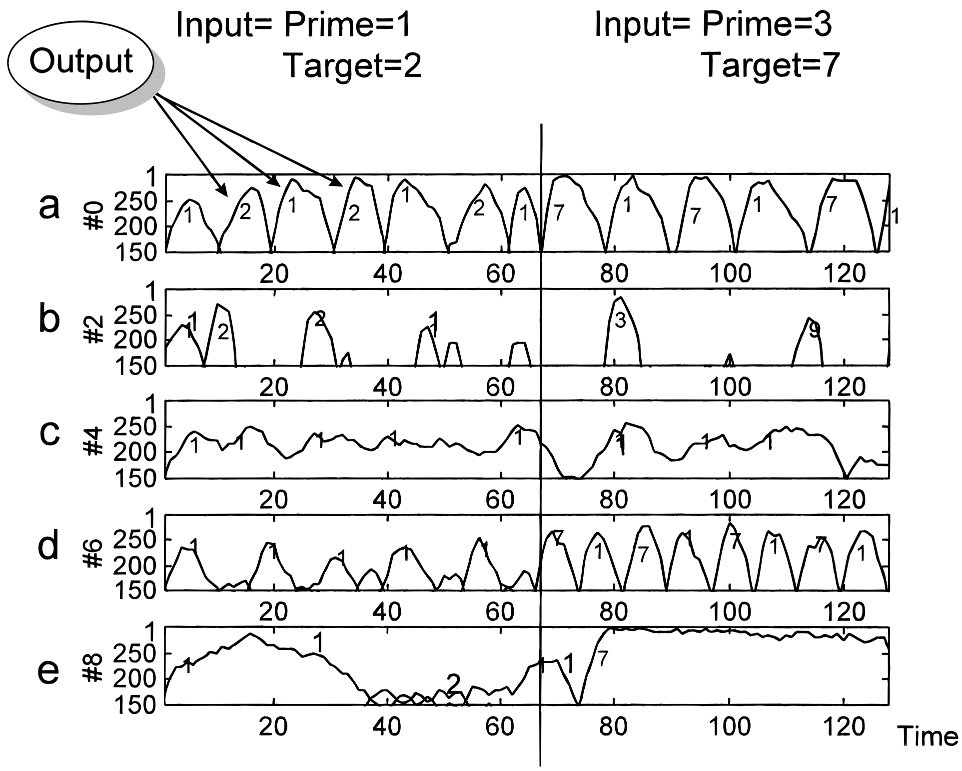

Fig. 5. (a) Simulation of normal priming (Memory Activation "1" simulates the prime and the Memory Activation "2" simulates the target concept). (b) Simulation of errors of lexical decisions. (c) Simulation of perseverative responses. (d,e) Simulation of fast and slow performances, respectively.

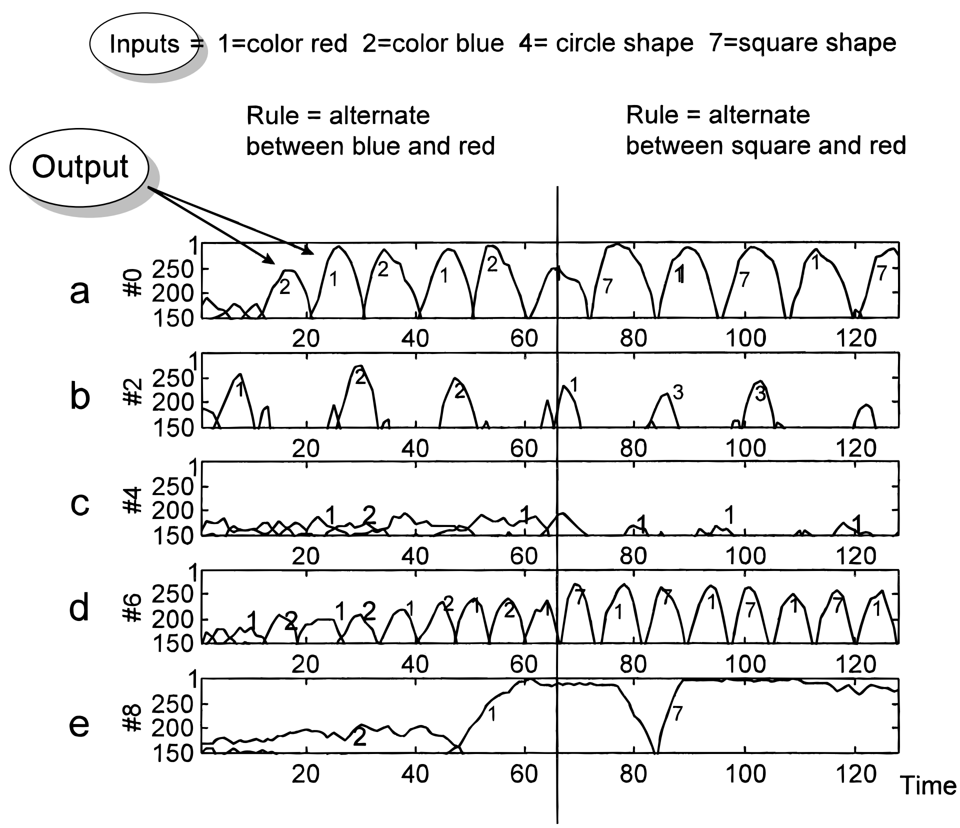

Fig. 6. (a) Simulation of "normal" object alternation task. (b) Simulation of errors of the object alternation task. (c) Simulation of perseverative responses. (d,e) Simulation of fast and slow performances, respectively.

https://doi.org/10.1017/S1355617700655108 Published online by Cambridge University Press

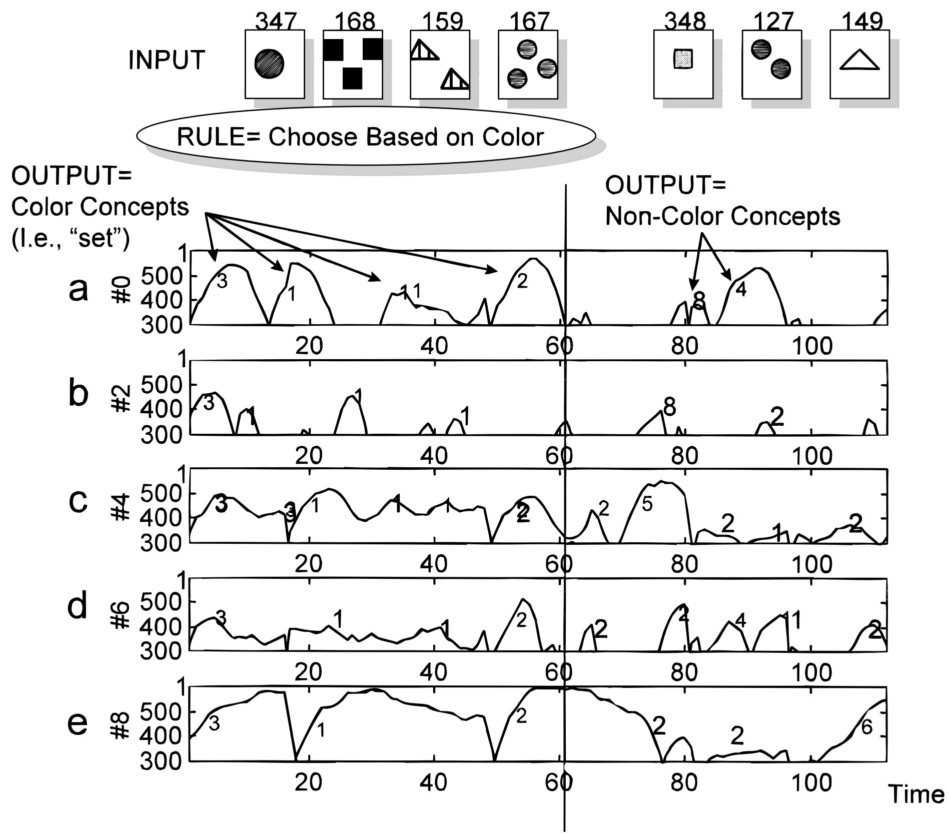

Fig. 7. The inputs are concrete stimuli of different colors and shapes. The card stimulus is simulated by a sum of input patterns from the different concrete concepts (e.g., the first card at the top of Figure 7a with *one green circle* will be a sum of memories *3*, *4*, and *7*). Activation of a color concept simulates a correct response to the sorting task (i.e., choosing according to color). The color principle (represented first by *3* and then by *1*) is activated when different input sets (cards) are presented. The color concept is stopped at a point marked by the vertical line. (a) Simulation of normal responses (color memories) are activated when the color principle is held in working memory and other different color concepts when the color principle is stopped. (b) Simulation of total failure on the task. (c) Tendency to repeat concepts *1* and *2* simulates the tendency to perseverate activating color concepts even though the internal input of the color principle was stopped. (d,e) Simulation of errors combined with different velocities.

is, memories *1* and *7* in (last 60 iterations of Figure 5a), may simulate indirect priming, that is, when the prime is related to the target via an additional associated concept (see introduction above). Note that in the latest case, where patterns are from different categories, the activation delay is longer than in the first case, where patterns are from the same categories.

In the object alteration task the presented stimuli are *yellow or red circle* and *red or yellow square*, which represented external inputs *1*, *2*, *7*, and *8* (Figure 6). The colors and shapes change randomly, that is, the circle may be blue in one iteration and red in the next iteration. In the first part of the simulated object alteration task, the rule is to alternate between yellow and blue colors and ignore the shapes, that is, alternate between memory *1* and memory *2*, represented by *1* and *2* in the internal input (Figure 6a). In the second part of the simulated object alteration task, the rule is to alternate between *red* and *square*, that is, alternate between memory *1* and memory *7*, represented by *1* and *7* in the internal input (Figure 6a); note that in this case the alteration rate is slower.

For the first part of the sorting task simulation (Figure 7 up to the vertical line), the internal input is set to a linear combination of memories *1*, *2*, and *3*, the first iterations, which represent the *color concept*, while different external input sets (cards) are presented. For the second part of the simulation, the *color concept* is stopped and additional cards are presented at the external input (after the vertical line, Figure 7). Simulation of normal responses activates *color memories* when the color principle is activated, while other different memories are activated when the color principle is stopped (Figure 7a). It is important to reemphasize that the *acquisition* of the abstract concept is not simulated; rather it is *artificially set up* by the internal input (Equation 5). As such, the model is limited to simulation of the actual maintenance of the principle once it has been acquired.

https://doi.org/10.1017/S1355617700655108 Published online by Cambridge University Press

Simulation of Pathological Cognitive Performances on the Tasks

As mentioned above, beyond the point of optimal sequence activity, the model simulates cognitive disturbances for the four different tasks. Panels *b*, *c*, *d*, and *e* in Figures 4, 5, 6, and 7 present typical "abnormal" results, each corresponding to a simulation with a different dynamic threshold combination according to Table 2. We choose to cluster together the simulation results of all the cognitive tasks for each point of dynamic threshold parameters in order to emphasize the similar "abnormalities" that occur in all the tasks.

Simulation *b*, the second simulation for all tasks (Figures 4b, 5b, 6b, and 7b), is achieved when the dynamic threshold parameters are set to the values in Row 2 of Table 2, namely, the point of high and fast threshold dynamics. Here the number of convergences is significantly reduced (i.e., increased spaces of no convergence between activations), and when a convergence of the network occurs, it is not always input dependent (i.e., a convergence to a memory that is not included in the input). This is most evident in the simulation of thinking (Figure 4b) where, in the first half of the simulations, there are very few partial activations, while in the second half of the same simulation, activations are not according to the inputs (i.e., memory *3* is activated even though only memories *1*, *4*, or *7* are presented). The same trend of reduced activations and errors also characterizes simulations of other tasks.

Simulation *c*, the third simulation for all tasks (Figures 4c, 5c, 6c, and 7c), is achieved when the dynamic threshold parameters are set to the values in Row 3 of Table 2, that is, the point of slow and low levels of threshold. Here, again, there are very few convergences at all, and when the network converges, it tends to be stuck in that state for the entire simulation; for example, in Figure 5c, memory number *1* is continuously correlated with the state of the system. When input is changed, the network responds with an attempt to activate another memory, but this is indicated by occasional activation of another memory or by fluctuation in correlations only. In the simulation of the sorting task activation of color memories tends to persist even after the sorting rule is stopped.

Simulation *d*, the fourth simulation for all tasks (Figures 4d, 5d, 6d, and 7d), is performed when the dynamic threshold parameters are set to the values in Row 4 of Table 2, that is, the point of fast but optimal levels of threshold. Here the simulations are dominated by significantly more activations or, in other words, faster activity. Furthermore, the activation of memories are mostly input dependent, for example, in the simulations of the sorting task and the object alternation tasks. In the sorting task there still persists a tendency to activate color concepts even when the rule to choose based on colors stops. Errors in *sequence* may occur for this threshold combination, it is evident in the thinking simulation where only *1* and *3* memories are reactivated in the first half of the simulation (Figure 4d).

Simulation *e*, the fifth simulation for all tasks (Figures 4e, 5e, 6e, and 7e), is performed when the dynamic threshold parameters are set to the values described in Row 5 of Table 2 (i.e., the point of slow but optimal levels of threshold). Strikingly the activity in the model is very slow; however, the convergences tend to be input dependent in most cases.

DISCUSSION

One of the more perplexing phenomena in the study of mental disorders (i.e., psychiatric signs and symptoms) is the difficulty in relating to the rich and variant spectrum of phenomena (Wilson, 1993; Tucker, 1998). Symptoms and cognitive disturbances appear mixed and in different combinations. This makes it difficult to create sets of distinct homogeneous disturbances, and even more, to relate these disturbances to any biological marker or hypothesized source for the symptoms (Van-Praag, 1997). The model in this work is useful only to the extent that it shows that a hypothetical origin of a mental disorder (in this case, the altered parameters of threshold dynamics) can be manifested by a large set of varying expressions, that probably lie in a spectrum range and not as distinct sets of disturbances.

Many recent studies (see, e.g., Cohen & Servan-Schreiber, 1992, 1993; Hermann et al., 1993; Hinton, 1981, 1986; Hinton & Shallice, 1991; Hoffman, 1987, 1992; Hoffman & Dobscha, 1989; Hoffman et al., 1994; Rumelhart & McClelland, 1986) suggest that memory activation into mental states that govern thought and speech might be simulated by the activity of an attractor NN model. In this study we employ a dynamic threshold semantic neural network (DTSNN) as a framework for simulation and performance analysis of a large variety of cognitive tasks in different conditions. By clustering the random memory patterns of the network into classes of neighboring patterns and assigning asymmetric connections between the units of the network that represent memory patterns of the same class, we have created a representation of a semantic network model. Each memory pattern can represent a concrete concept or a node in the semantic network, whereas each class of associated concrete concepts can represent an abstract concept. The model can be further refined by using the degree of membership of each concept in its associated class or classes to set the strength of its asymmetric connections to other concepts in its class. By adding an internal input to the network we naturally represent a short-term working memory. The dynamics processes of acquiring this working memory should be further investigated. One possible way is that any new meaningful stimulus (one that causes the system to converge), will be kept for a short time as part of the internal input, and that the linear combination of these short-term memories would form an internal rule that may govern the system behavior.

The simulations show that by varying the dynamic parameters of the threshold large and rich sets of disturbances can be obtained in the activity of the model. Slower and faster activity, errors, and persistent convergences are some of these disturbances. The disturbances can be viewed as lying on a continuum. For example, it has been demonstrated that gradual changes in threshold velocity rates correspond to gradual changes in convergence frequency (i.e., faster or slower memory activations). Errors typically occur when the threshold combination is not optimal; however, errors are not related to convergence velocity, and they can occur in slow or in fast memory activations. Errors can be in the *order* of memory activation or in the *type* of memory activation. When a memory is activated, even though it does not correspond to any of the input patterns, it is then defined as *non-input-dependent* memory activation. Persistence of activity can be the case where the model activates the same memory continuously, but it can also be the case when the previous memory is activated over and over again even though the input has already been changed.

These mixtures of different disturbances appear to various extents in the simulations of abnormal conditions. They are rough simulations of clinical descriptions from patients that suffer from a variety of mental disorders. It is emphasized that these are merely metaphorical resemblances and that quantification and correlations to the specific task data are required for a more specific simulation of the tasks. Metaphorically, alterations in the velocity of memory activations may simulate alterations in the velocity of ideation or speech: for example, the symptoms related to *flight of ideas* or *pressure of speech* (Andreasen, 1997; Ariety & Goldstein, 1959). Errors involving order of memory activations and of non-input-dependent memory activations may simulate thought disorders such as loosening of associations and delusions. Loosening of association can be the case where the goal-directed sequences are disrupted by memories activated not according to the relevant sequence. When a memory is activated regardless of input (i.e., non-input-memory activation) it may simulate an idea or thought that does not have any roots in the events of the real world: for example, the idea of being persecuted even though nothing is really happening to suggest threat. Persistence of the same memory or reactivation of previous memories even though new ones are presented may simulate poverty of thought and perseverations (Andreasen & Olsen, 1982). When the different tasks are considered, the failures have been attributed to a variety of reasons, among them the tendency to perseverate and respond to previous stimuli (Andreasen, 1983).

As emphasized above, quantification and correlations of simulation results to the specific task data is required for a more authentic simulation of the tasks. However, even at the metaphorical phase of this work, the model is useful for predicting a spectrum of varying patterns of disorders rather than sets of homogeneous error types. In other words the model both describes and predicts that the relations between an original hypothetical mechanism of neural pathology and its resulting disturbances do not have a one-to-one correlation, but rather a spectrum or range of alterations.

The dynamic threshold parameters were chosen as the hypothetical neural mechanism that is generating the disturbances. This choice came out of the realization that the threshold dynamics have a crucial effect on the organizational activity of the model (regarding ordered convergences and patterns of memory activations; Horn & Usher, 1989, 1990). Interestingly, the importance of threshold dynamics can be found in recent neuroscience literature (Fuster, 1997; Globus, 1992; King, 1991; Roland, 1993; Rumelhart & McClelland, 1986). Indirectly, threshold may relate also to another aspect of brain research, that of connectivity. Increased thresholds cut off inputs from outputs in the units of the model by deactivating each unit. Although this is not an accurate description of a disconnection syndrome (Friston, 1996), with some imagination one may view "input–output disconnection" as a kind of disconnection. The emphasis on the role of threshold activity in mental disorders has been notably described by Cohen and Servan-Schreiber (1992) and other work by Cohen (Cohen & Servan-Schreiber, 1993). The disconnection theory developed by Friston and associates (1996, 1998; Friston & Frith, 1995; Friston et al., 1994; Frith, 1992) for mental disturbances, especially in schizophrenia is a leading approach to understanding this mental disorder. Fast and high threshold dynamics in the model disconnect the input from the output in the model, and this results in errors and jumps (Figures 4b, 5b, 6b, and 7b), which can simulate thought disorders in schizophrenia.

The model simulation results may offer an objective classification system of various thought disorders by their typical errors. There are still many points in the threshold space that show interesting disturbances. The model may help to design cognitive tests for more careful classification and diagnosis of the various thought disorders. One may think of some continuous *trajectories in the dynamic-threshold parameters space* (Figure 3) that can be relevant for describing the *course of some mental diseases*, such as schizophrenia. Loosening of associations is more typical of the acute psychotic phases of the disease, while poverty of thought content more frequently characterizes the postpsychotic, residual stages of the disease. Schizophrenia progresses from one psychotic phase to another, and accordingly a progressive degradation of mental ability with poverty of thought content occurs. The relation between the disturbances in the model and the threshold parameters may broaden our insight into what such parameters, or their derivatives, might be in biological systems.

In summary, even though the model is highly theoretical and metaphoric, if threshold dynamics in the model can help us to gain certain insights into the relation between thresholds or connectivity and clinical symptoms in real brains of patients, then the model can be useful as a theoretical tool for further research.

REFERENCES

Abbott, L.F., Varela, J.A., Sen, K., & Nelson, S.B. (1997). Synaptic depression and cortical gain control. *Science*, *275*, 220–224.

Andreasen, N.C. (1983). *The scale for the assessment of negative symptoms (SANS)*. Iowa City: University of Iowa.

https://doi.org/10.1017/S1355617700655108 Published online by Cambridge University Press

Andreasen, N.C. (1997). Linking mind and brain in the study of mental illnesses: A project for a scientific psychopathology. *Science*, *275*, 1586–1596.

Andreasen, N.C. & Olsen, S. (1982). Negative and positive schizophrenia: Definition and validation. *Archives of General Psychiatry, 39*, 789–794.

Berman, K.F., Osterm, J.L., & Weinberger, D.R. (1995). Physiological activation of a cortical network during performance of the Wisconsin card sorting test: A positron emission tomography study. *Neuropsychologia*, *33*, 1027–1046.

Bliss, T.V.P. & Gardner-Medwin, A.R. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the unanaesthetised rabbit following stimulation of the prefrontal path. *Journal of Physiology*, *232*, 357–374.

Cohen, J.D. & Servan-Schreiber, D. (1992). Context, cortex, and dopamine: A connectionist approach to behavior and biology in schizophrenia. *Psychological Review*, *82*, 407–428.

Cohen, J.D. & Servan-Schreiber, D. (1993). A theory of dopamine function and its role in cognitive deficits in schizophrenia. *Schizophrenia Bulletin*, *19*, 85–103.

Collins, A.M. & Loftus, E.F. (1975). A spreading activation theory of semantic processing. *Psychological Review*, *82*, 407–428.

Collins, A.M. & Quillian, M.R. (1970). Facilitating retrieval from semantic memory: The effect of repeating part of an inference. *Acta Psychologica*, *33*, 304–314.

Desse, J. (1987). Thought into speech. *Scientific American*, *66*, 314–321.

Friston, K.J. (1996). Theoretical neurobiology and schizophrenia. *British Medical Bulletin*, *52*, 644–655.

Friston, K.J. (1998). *Context-sensitive interactions in the brain*. Paper presented at the Cognitive Neuroscience Society Annual Meeting, San Francisco.

Friston, K.J. & Frith, C.D. (1995). Schizophrenia a disconnection syndrome? *Clinical Neuroscience*, *3*, 89–97.

Friston, K.J., Liddel, P.F., & Frith, C.D. (1994). Dysfunctional fronto–temporal integration in schizophrenia. *Neuropsychopharmacology*, *10*, 538s.

Frith, C.D. (1992). *The cognitive neuropsychology of schizophrenia*. Hove, UK: Lawrence Erlbaum.

Fuster, J.M. (1997). Network memory. *Trends in Neuroscience*, *20*, 451–459.

Globus, G. (1992). Toward a noncomputational cognitive neuroscience. *Journal of Cognitive Neuroscience*, *4*, 299–310.

Goldman-Rakic, P.S. (1994). Working memory dysfunction in schizophrenia. *Journal of Neuropsychiatry*, *6*, 348–356.

Goldstein, K. (1959). *Organismic Approach*. In Ariety, S. (Ed.), pp. 1333–1337. New York: Basic Books.

Hermann, M., Ruppin, E., & Usher, M. (1993). A neural model for dynamic activation of memory. *Biological Cybernetics*, *68*, 455–463.

Herz, J., Krogh, A., & Richard, G.P. (1991). *Introduction to the theory of neural computation*. Santa Fe, NM: Santa Fe Institute–Addison Wesley.

Hinton, G.E. (1981). *Implementing semantic networks in parallel hardware*. Hillsdale, NJ: Erlbaum.

Hinton, G.E. (1986). *Learning distributed representations of concepts*. Hillsdale, NJ: Erlbaum.

Hinton, G.E. & Shallice, T. (1991). Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review*, *98*, 74–95.

Hoffman, R.E. (1987). Computer simulation of neural information processing and the schizophrenia mania dichotomy. *Archive of General Psychiatry*, *44*, 178–187.

Hoffman, R.E. (1992). Attractor neural networks and psychotic disorders. *Psychiatric Annals*, *22*, 119–124.

Hoffman, R.E. & Dobscha, S. (1989). Cortical pruning and the development of schizophrenia: A computer model. *Schizophrenia Bulletin*, *15*, 477–490.

Hoffman, R.E., Oats, E., Hafner, J., & Husting, H.H. (1994). Semantic organization of hallucinated "voices" in schizophrenia. *American Journal of Psychiatry*, *151*, 1229–1230.

Horn, D. & Usher, M. (1989). Neural networks with dynamic thresholds. *Physics Review*, *40*, 1036–1044.

Horn, D. & Usher, M. (1990). Excitatory inhibitory networks with dynamic thresholds. *International Journal of Neural Systems*, *1*, 249–257.

King, C.C. (1991). Fractal and chaotic dynamics in nervous systems. *Progress in Neurobiology*, *36*, 279–308.

Maher, B.A., Manschreck, T.C., & Molino, M.A.C. (1983). Redundancy pause distribution and thought disorder in schizophrenia. *Language and Speech*, *26*, 191–199.

Manschreck, T.C., Maher, B.A., & Milavetz, J.J. (1988). Semantic priming in thought-disordered schizophrenic patients. *Schizophrenia Research*, *1*, 61–66.

McNaughton, B.L. (1982). Long-term synaptic enhancement and short-term potentiation in rat facia dentata VTA through different mechanisms. *Journal of Physiology*, *18*, 249–262.

Neely, J.H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited capacity attention. *Journal of Experimental Psychology*, *106*, 226–254.

Neely, J.H. (1991). *Semantic priming effects in visual word recognition: A selective review of current findings and theories*. Hillsdale, NJ: Erlbaum.

Roland, P.E. (1993). *Brain activation*. Stockholm: Wily-Liss Inc.

Rumelhart, D.E. & McClelland, J.L. (1986). *Parallel distributed processing: Exploration in the microstructure of cognition*. (PDP Reaserch Group ed., Vol. 1 and 2). Cambridge, MA: MIT Press.

Servan-Schreiber, D., Cohen, J.D., & Steingard, S. (1996). Schizophrenia deficits in the processing of context: A test of a theoretical model. *Archives of General Psychiatry*, *53*, 1105–1112.

Spitzer, R.L. & Williams, J.B.W. (1995). *Structural Clinical Interview for DSM IV, Patient version*. New York: New York State Psychiatric Institute.

Stone, D.J., Dozovsky, I., Morgan, T.E., Anderson, C.P., & Finch, C.E. (1998). Increased synaptic sprouting in response to estrogen via apolipoprotein E-dependent mechanism: Implication for Alzheimer's disease. *Journal of Neuroscience*, *18*, 3180–3185.

Tucker, G.J. (1998). Putting DSM–IV in perspective. *American Journal of Psychiatry*, *155*, 159–161.

Van-Praag, H.M. (1997). The future of biological psychiatry. *CNS Spectrums*, *2*, 18–25.

Wickliffe, A.C. & Warren, T.P. (1997). Metaplasticity: A new vista across the field of synaptic plasticity. *Progress in Neurobiology*, *52*, 303–323.

Wilson, M. (1993). DSM–III and the transformation of American psychiatry: A history. *American Journal of Psychiatry*, *3*, 399–409.